

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 12:23:23 ; Search time 29 Seconds
(without alignments)
92.366 Million cell updates/sec

Title: US-09-632-429-4

Perfect score: 89

Sequence: 1 WEVLCTWETCER 13

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriapi:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	52	58.4	1858	12 Q91PR6	Q91pr6 squash mosa
2	51	57.3	292	3 Q05533	Q05533 saccharomyc
3	49	55.1	395	10 Q9C693	Q9c693 arabidopsis
4	49	55.1	533	10 Q8RY80	Q8ry80 arabidopsis
5	47	52.8	236	13 P79899	P79899 oncorhynch
6	47	52.8	236	13 P79905	P79905 salmo salar
7	46	51.7	173	12 Q8QLJ0	Q8qlj0 mamestra co
8	46	51.7	427	3 O74498	O74498 schizosacch
9	46	51.7	1042	5 Q9GRV9	Q9grv9 caenorhabdi
10	46	51.7	1080	5 O01979	O01979 caenorhabdi
11	46	51.7	1121	5 O02252	O02252 caenorhabdi
12	46	51.7	1230	2 Q59325	Q59325 clostridium
13	45.5	51.1	808	5 Q9V995	Q9v995 drosophila
14	45	50.6	426	2 O87830	O87830 streptomyce
15	45	50.6	895	2 O68438	O68438 clostridium
16	44	49.4	155	17 Q8TNN4	Q8tnm4 methanosarc

17	44	49.4	168	13 Q9DF06	Q9dfq6 gillichthys
18	44	49.4	350	17 Q8TVT5	Q8tvt5 methanopyru
19	43	48.3	205	4 Q8T57	Q8t57 homo sapien
20	43	48.3	361	2 Q8RSU9	Q8rsu9 corynebacte
21	43	48.3	394	5 Q8TIA3	Q8tia3 trypanosoma
22	43	48.3	516	10 Q42701	Q42701 catharanthu
23	43	48.3	524	10 Q42700	Q42700 catharanthu
24	42.5	47.8	239	2 Q9ZIU0	Q9ziu0 burkholderi
25	42.5	47.8	239	2 Q9AM50	Q9am50 burkholderi
26	42.5	47.8	239	2 Q9AM49	Q9am49 burkholderi
27	42.5	47.8	239	2 Q9AM48	Q9am48 burkholderi
28	42.5	47.8	239	2 Q9AM47	Q9am47 burkholderi
29	42.5	47.8	239	2 Q9AM46	Q9am46 burkholderi
30	42.5	47.8	239	2 Q93PF9	Q93pf9 burkholderi
31	42.5	47.8	239	2 Q93PF7	Q93pf7 burkholderi
32	42.5	47.8	239	2 Q93PF6	Q93pf6 burkholderi
33	42.5	47.8	239	2 Q93PF4	Q93pf4 burkholderi
34	42.5	47.8	239	2 Q93NN8	Q93nn8 burkholderi
35	42.5	47.8	239	2 Q9AM51	Q9am51 burkholderi
36	42.5	47.8	1405	11 Q8VHS2	Q8vhs2 mus musculu
37	42	47.2	347	17 Q29274	Q29274 archaeoglob
38	42	47.2	358	10 Q9SBS0	Q9sbs0 medicago va
39	42	47.2	426	4 Q8WXG7	Q8wxg7 homo sapien
40	42	47.2	475	10 Q9SD14	Q9sd14 arabidopsis
41	42	47.2	478	5 Q95YS8	Q95ys8 leishmania
42	42	47.2	501	10 Q9SYJ6	Q9syj6 arabidopsis
43	42	47.2	502	10 Q93WC5	Q93wc5 arabidopsis
44	42	47.2	545	13 Q90ZP5	Q90zps oncorhynch
45	42	47.2	553	13 Q9PTS9	Q9pts9 gallus gall

ALIGNMENTS

RESULT 1

Q91PR6 PRELIMINARY; PRT; 1858 AA.
ID Q91PR6
AC Q91PR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polyprotein.
OS Squash mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Comovirus.
OX NCBI_taxid=12263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y-SQMV;
RA Iwanami T., Han S.S., Karasev A.V.;
RT "Nucleotide sequence of a Japanese isolate of Squash mosaic virus.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB054688; BAB62139.1; -.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
SQ SEQUENCE 1858 AA; 209972 MW; B344D1FC83641829 CRC64;

Query Match 58.4%; Score 52; DB 12; Length 1858;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 WEVLCTWTE 9

DB 1033 WDFVCFDWE 1041

RESULT 2

Q05533 PRELIMINARY; PRT; 292 AA.
ID Q05533
AC Q05533;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

```

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to inositol monophosphatase.
GN YDR287W OR D9819.7.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Fulton L.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA Wilson R., Waterston R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Waterston R.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Jia Y., Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U51031; AAB64472.1; -.
DR HSP; P29218; IMF.
DR SGD; S0002695; YDR287W.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; UNKNOWN_1.
SQ SEQUENCE 292 AA; 37092 MW; CEA9D943F69E2082 CRC64;

Query Match 57.3%; Score 51; DB 3; Length 292;
Best Local Similarity 54.5%; Pred. No. 3.6;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 WEVLCWTWETC 11
|| || | |
Db 223 WEGGCWANDVC 233

RESULT 3
Q9C693
ID Q9C693 PRELIMINARY; PRT; 395 AA.
AC Q9C693;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 46.3 kDa protein.
GN t8L23.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

```

```

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
RA Militischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer J.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT *Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana*.
RL Nature 408:816-820(2000).
DR EMBL; AC079733; AAG50749.1; -.
DR InterPro; IPR001202; WW_RSP5_WMP.
DR PROSITE; PS01159; WW_DOMAIN_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 395 AA; 46301 MW; B798F3466E3B62D3 CRC64;

Query Match 55.1%; Score 49; DB 10; Length 395;
Best Local Similarity 54.5%; Pred. No. 9.4;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 WEVLCWTWETC 11
|::| | | | |
Db 284 WKLLSWAWLTC 294

RESULT 4
Q8RY80
ID Q8RY80 PRELIMINARY; PRT; 533 AA.
AC Q8RY80;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 62.8 kDa protein.
GN ATIG57600.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Ouach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT *Arabidopsis Full Length cDNA Clones.*;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074538; AAL69506.1; -.
KW Hypothetical protein.
SQ SEQUENCE 533 AA; 62791 MW; 5C497EF542EB9A66 CRC64;

Query Match 55.1%; Score 49; DB 10; Length 533;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 WEVLCWTWETC 11
|::| | | | |
Db 422 WKLLSWAWLTC 432

RESULT 5
P79899
ID P79899 PRELIMINARY; PRT; 236 AA.
AC P79899;

```

```
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Pentraxin precursor.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RX MEDLINE=97131713; PubMed=8977214;
RA Jensen L.E., Hiney M.P., Shields D.C., Uhlar C.M., Lindsey A.J.,
RA Whitehead A.S.;
RT "Acute phase proteins in Salmonids. Evolutionary analyses and acute
phase response.";
RL J. Immunol. 158:384-392(1997).
DR EMBL; X99385; CAA67764.1; -.
DR HSSP; P02743; 1SAC.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF00354; pentaxin; 1.
DR PRINTS; PR00895; PENTAXIN.
DR ProDom; PD002153; Pentaxin; 1.
DR SMART; SM00159; PTX; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT CHAIN.
SQ SEQUENCE 236 AA; 26835 MW; 7F39A5F559025857 CRC64;

Query Match 52.8%; Score 47; DB 13; Length 236;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 WEVLCWTWET 10
DB 116 WISICWTWDS 125

RESULT 6
ID P79905 PRELIMINARY; PRT; 236 AA.
AC P79905;
DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Pentraxin precursor.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RX MEDLINE=97131713; PubMed=8977214;
RA Jensen L.E., Hiney M.P., Shields D.C., Uhlar C.M., Lindsey A.J.,
RA Whitehead A.S.;
RT "Acute phase proteins in Salmonids. Evolutionary analyses and acute
phase response.";
RL J. Immunol. 158:384-392(1997).
DR EMBL; X99386; CAA67765.1; -.
DR HSSP; P02743; 1SAC.
DR InterPro; IPR001759; Pentaxin.
DR Pfam; PF00354; pentaxin; 1.
DR PRINTS; PR00895; PENTAXIN.
DR ProDom; PD002153; Pentaxin; 1.
DR SMART; SM00159; PTX; 1.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT CHAIN.
SQ SEQUENCE 236 AA; 26786 MW; FA69D8A65A5B9BE7 CRC64;
```

```
Query Match 52.8%; Score 47; DB 13; Length 236;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 WEVLCWTWET 10
DB 116 WISICWTWDS 125

RESULT 7
ID Q8QLJ0 PRELIMINARY; PRT; 173 AA.
AC Q8QLJ0;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 20.2 kDa protein.
OS Mamestra configurata nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=191492;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=90/2;
RX MEDLINE=97163493; PubMed=9010313;
RA Li S., Erlandson M., Moody D., Gillott C.;
RT "A physical map of the Mamestra configurata nucleopolyhedrovirus
genome and sequence analysis of the polyhedrin gene.";
RL J. Gen. Virol. 78:265-271(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=90/2;
RX MEDLINE=21884635; PubMed=11886270;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.;
RT "Sequence and Organization of the Mamestra configurata
Nucleopolyhedrovirus Genome.";
RL Virology 294:106-121(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=90/2;
RA Li Q., Donly C., Li L., Willis L.G., Theilmann D.A., Erlandson M.A.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
KW Hypothetical protein.
SQ SEQUENCE 173 AA; 20238 MW; F85DFDAA70912360 CRC64;

Query Match 51.7%; Score 46; DB 12; Length 173;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLCWTWETCER 13
DB 146 VRCWVWGLCEK 156

RESULT 8
ID O74498 PRELIMINARY; PRT; 427 AA.
AC O74498;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Hypothetical 48.7 kDa protein C285.11 in chromosome III.
GN SCC285.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RA Seeger K., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
```

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO HUMAN KIAA0887.
 DR EMBL; AL031545; CAA20850.1; -.
 DR InterPro; IPR001012; UBX.
 DR Pfam; PF00789; UBX; 1.
 DR SMART; SM00166; UBX; 1.
 KW Hypothetical protein; Coiled coil.
 FT DOMAIN 248 315 COILED COIL (POTENTIAL).
 SQ SEQUENCE 427 AA; 48670 MW; 3AA7E161BE1F7D90 CRC64;

Query Match 51.7%; Score 46; DB 3; Length 427;
 Best Local Similarity 60.0%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VLTWTWETCE 12
 :||||: ||
 Db 178 ILCTWGDVCE 187

RESULT 9

Q9GRV9 PRELIMINARY; PRT; 1042 AA.
 AC Q9GRV9;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Y116F11B.3 protein.
 GN Y116F11B.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sulston J.E., McLay K.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]

QY SEQUENCE FROM N.A.
 ID MEDLINE-99069613; PubMed=9851916;
 AC none;
 DT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; ALI32943; CAC14390.1; -.
 SQ SEQUENCE 1042 AA; 116388 MW; FCD2A79BD7359B60 CRC64;

Query Match 51.7%; Score 46; DB 5; Length 1042;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCWTWETC 11
 :||||:
 Db 854 LLWTWQTC 861

RESULT 10

Q01979 PRELIMINARY; PRT; 1080 AA.
 ID Q01979; O02253;
 AC Q01979 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 20, Last annotation update)
 DE F23B2.12 protein.
 GN F23B2.12.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE-94150718; PubMed=7906398;
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA McMurray A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z82266; CAB05187.1; -.
 DR EMBL; Z68295; CAB05187.1; JOINED.
 DR EMBL; Z68295; CAA92588.1; -.
 DR EMBL; Z82266; CAA92588.1; JOINED.
 DR InterPro; IPR000379; Ser_estrs_site.
 SQ SEQUENCE 1080 AA; 121535 MW; C34E65D19BA7782 CRC64;

Query Match 51.7%; Score 46; DB 5; Length 1080;
 Best Local Similarity 75.0%; Pred. No. 67;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCWTWETC 11
 :||||:
 Db 886 LLWTWQTC 893

RESULT 11

O02252 PRELIMINARY; PRT; 1121 AA.
 ID O02252
 AC O02252;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE F23B2.11 protein.
 GN F23B2.11.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z82266; CAB05185.1; -.
 DR InterPro; IPR000073; Abhydrolase.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF00561; abhydrolase; 1.
 SQ SEQUENCE 1121 AA; 125951 MW; A1B4D7BD92116EB5 CRC64;

Query Match 51.7%; Score 46; DB 5; Length 1121;
 Best Local Similarity 75.0%; Pred. No. 69;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCWTWETC 11
 :||||:
 Db 928 LLWTWQTC 935

```

RESULT 12
Q59325
ID Q59325 PRELIMINARY; PRT; 1230 AA.
AC Q59325;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cellulose 1,4-beta-CELLULOSIDASE (EC 3.2.1.91) (Exoglucanase)
DE (Exocellobiohydrolase) (1,4-beta-cellobiohydrolase).
GN CBH3 OR CBHA
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F7;
RA Zverlov V.V.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-315 FROM N.A.
RC STRAIN=F7;
RA Zverlov V.V.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES
CC IN CELLULOSE AND CELLOTETRAOSE, RELEASING CELLOBIOSE FROM THE NON-
CC REDUCING ENDS OF THE CHAINS.
DR EMBL; X80993; CAA56918.1; -.
DR EMBL; AJ005783; CAA06693.1; -.
DR HSSP; Q06851; INNC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR003305; CBM_GenC.
DR InterPro; IPR004197; celd_N.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001701; GH_9.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF02018; CBM_4_9; 1.
DR Pfam; PF02927; celd_N; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR ProDom; PD001947; CBD_3; 1.
DR PROSITE; PS00448; CLOS_CELLULOSE_RPT; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1230 AA; 138077 MW; A398D9814B5D6A0E CRC64;

Query Match 51.7%; Score 46; DB 2; Length 1230;
Best Local Similarity 63.6%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVLCWTWETCE 12
DB 50 EGLCYPWHTCE 60

RESULT 13
Q9V995
ID Q9V995 PRELIMINARY; PRT; 808 AA.
AC Q9V995;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG15236 protein.
GN CG15236
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Query Match 51.1%; Score 45.5; DB 5; Length 808;
Best Local Similarity 50.0%; Pred. No. 61;
Matches 6; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 EVLCWTWETCER 13
DB 131 EIFCW-WEKCDK 141

RESULT 14
OB7830
ID OB7830 PRELIMINARY; PRT; 426 AA.
AC OB7830;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Glycosyltransferase.
GN OLEG1.
OS Streptomyces antibioticus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11891;

```

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA Ananides P.G., Scherer S.E., Ashburner M., Henderson S.N.,

RA George R.A., Lewis S.E., Richards S., Zhang Q., Chen L.X.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brannon R.C., Rogers J.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003790; AAF57401.1; -.

DR FlyBase; FBgn0033108; CG15236.

SQ SEQUENCE 808 AA; 87716 MW; 04AC2B699DA5F9AD CRC64;

Query Match 51.1%; Score 45.5; DB 5; Length 808;

Best Local Similarity 50.0%; Pred. No. 61;

Matches 6; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 EVLCWTWETCER 13

DB 131 EIFCW-WEKCDK 141

RESULT 14

OB7830

ID OB7830 PRELIMINARY; PRT; 426 AA.

AC OB7830;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

DE Glycosyltransferase.

GN OLEG1.

OS Streptomyces antibioticus.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1890;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 11891;

RX MEDLINE=98420102; PubMed=9749673;
RA Olano C., Rodriguez A.M., Michel J.M., Mendez C., Raynal M.C.,
RA Salas J.A.;
RT "Analysis of a Streptomyces antibioticus chromosomal region involved
RT in oleandomycin biosynthesis, which encodes two glycosyltransferases
RL Mol. Gen. Genet. 259:299-308(1998).
DR EMBL; AJ02638; CAA05641.1; -
KW Transferase.
SQ SEQUENCE 426 AA; 47008 MW; 4BADDD0551BC25EC CRC64;

Query Match 50.6%; Score 45; DB 2; Length 426;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 WTWETCER 13
Db 239 WWPCEER 246

RESULT 15
O68438
ID O68438 PRELIMINARY; PRT; 895 AA.
AC O68438;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
DE Cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91).
GN CELK.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JW 20;
RX MEDLINE=99395035; PubMed=10464199;
RA Kataeva I., Li X.L., Chen H., Choi S.K., Ljungdahl L.G.;
RT "Cloning and sequence analysis of a new cellulase gene encoding CelK,
RT a major cellulosome component of Clostridium thermocellum: evidence
RT for gene duplication and recombination.";
RL J. Bacteriol. 181:5288-5295(1999).
DR EMBL; AF039030; AAC06139.1; -.
DR HSSP; P04954; ICLC.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR004197; celd_N.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001701; GH_9.
DR Pfam; PF02018; CBM_4_9; 1.
DR Pfam; PF02927; celd_N; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PROSITE; PS00448; CLOS_CELLOLOSOME_RPT; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 895 AA; 100712 MW; 5DB1FD84A6750CCE CRC64;

Query Match 50.6%; Score 45; DB 2; Length 895;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVLCWTWETCE 12
Db 50 EGLCFPWHTCE 60

Search completed: January 9, 2003, 12:25:45
Job time : 30 secs